

## Narzędzia cyfrowe w polonistycznej dydaktyce akademickiej – zastosowania, możliwości, perspektywy

Słowa klucze: humanistyka cyfrowa, lingwistyka cyfrowa, narzędzia cyfrowe, dydaktyka akademicka

Zwrot cyfrowy w humanistyce nie jest już dziś niczym nowym. Od wielu lat, również w Polsce, możemy obserwować powstawanie interdyscyplinarnych zespołów badawczych (skupiających informatyków i przedstawicieli szeroko pojętej humanistyki), rozwój badań, w których wykorzystywane są narzędzia cyfrowe, a także wymianę myśli i doświadczeń badaczy podczas konferencji i seminariów poświęconych cyfrowości w różnych obszarach humanistyki. Radosław Bomba zwraca uwagę na fakt, że:

O ile do tej pory humanistyka w zasadniczej mierze opierała się na tekście, o tyle obecnie coraz większą rolę w przeprowadzaniu badań zaczynają odgrywać cyfrowe instrumenty. Jest to sytuacja, która nie występowała wcześniej na taką skalę. W dotychczasowych badaniach humanistycznych najczęściej mieliśmy do czynienia z sytuacją, w której twierdzenie, wynik badań, artykuł naukowy, opierały się na innych tekstach, a nie na wynikach dostarczanych przez narzędzia i instrumenty.

BOMBA, 2013: 64

Wydaje się więc, że trudno dziś uniknąć w humanistyce związków z cyfrowością. Warto zatem rozważyć, czy, a jeśli tak, to w jaki sposób, można wykorzystać narzędzia z obszaru lingwistyki cyfrowej<sup>1</sup> w nauczaniu akademickim, ściślej: w dydaktyce przedmiotów językoznawczych. Być może jest to nawet konieczne, biorąc pod uwagę sylwetkę współczesnego studenta. Opierając się na przywołanym przez Błażeja Sajduka podziale na cyfrowych imigrantów i cyfrowych tubylców, współcześni studenci należą do tej drugiej

---

<sup>1</sup> Lingwistykę cyfrową rozumiemy zgodnie z definicją zaproponowaną przez twórców projektu DigiLing – Trans-European eLearning Hub for Digital Linguistics: „What is Digital Linguistics? It is a new interdisciplinary field which develops knowledge and skills for understanding, processing and exploiting language content in the digital era. This new discipline is a fusion between linguistics and computing, creating powerful tools at the crossroads of traditional language studies and information technologies” (DigiLing: online; data dostępu: 21.11.2017). [„Czym jest lingwistyka cyfrowa? Jest to nowa interdyscyplinarna dziedzina, rozwijająca wiedzę i umiejętności potrzebne do zrozumienia, przetwarzania i wykorzystywania treści językowych w erze cyfrowej. Ta nowa dyscyplina, będąc połączeniem lingwistyki i informatyki, tworzy potężne narzędzia na skrzyżowaniu tradycyjnych badań językowych i technologii informacyjnych”. – tłum. B.D., K.L.].

grupy: „[...] wyrosli razem z internetem, świat bez sieci znają tylko z opowiadań. Ich naturalnym habitatem jest świat wirtualny, do którego wchodzi nie tylko poprzez komputery, ale również za pośrednictwem tabletów lub/i telefonów komórkowych” (SAJDUK, 2014: 13). Współczesny student jest więc jednostką, która często korzysta z dorobku świata cyfrowego. Wiele jego aktywności związanych jest właśnie z przestrzenią nowoczesnych technologii, warto zatem wykorzystać ten potencjał również w obszarze dydaktyki. Sądzymy, że jest to istotne, ponieważ – jak zauważają Ewa Biłas-Pleszak i Katarzyna Sujkowska-Sobisz – „[...] obserwując współczesnych kandydatów na studia, czujemy, że musimy zmienić sposób myślenia o dydaktyce akademickiej, szukać nowych rozwiązań, dyskutować o rodzajach zmiany i sytuowania samych siebie w procesie edukowania” (BIŁAS-PLESZAK, SUJKOWSKA-SOBISZ, 2016: 168).

Cel w naszym artykule stanowi zaprezentowanie wybranych narzędzi oraz aplikacji internetowych, które mogą okazać się cennym wsparciem w dydaktyce akademickiej i pozwolią uatrakcyjnić zajęcia kursowe skierowane do studentów polonistyki. Nie sposób omówić w tym miejscu wszystkich dostępnych zasobów – nasz wybór jest więc w pełni subiektywny, podyktowany doświadczeniami dydaktycznymi, a także rozmowami przeprowadzonymi ze studentami na temat ich oczekiwań w zakresie uobecnienia lingwistyki cyfrowej na zajęciach. Świadomie pomijamy obszerne zasoby elektroniczne z zakresu leksykografii, zarówno współczesnej, jak i dawnej<sup>2</sup>. Jednocześnie zdajemy sobie sprawę z tego, że w wielu krajowych ośrodkach akademickich prowadzone są odrębne zajęcia poświęcone na przykład lingwistyce korpusowej. Nasze propozycje nie wyczerpują zatem tematu, wierzymy jednak, że mogą stanowić inspirację do przygotowania zajęć kursowych z wykorzystaniem nowoczesnych technologii i narzędzi humanistyki cyfrowej.

## Narodowy Korpus Języka Polskiego (NKJP)

Prace nad Narodowym Korpusem Języka Polskiego (<http://nkjp.pl>) zostały zakończone w 2012 roku, jednak niezmiennie pozostaje on najobszerniejszym anotowanym morfologicznie zbiorem danych języka polskiego<sup>3</sup>. Jego prymat w rodzimej lingwistyce korpusowej sprawia, że jest on nieocenionym źródłem materiału badawczego dla przedstawicieli wielu dyscyplin językoznawczych. Warto jednak zauważyć, że – analogicznie do różnych pod względem metodologicznym i tematycznym badań, prowadzonych z wykorzystaniem zawartych w nim danych – korpus ogólny polszczyzny może stanowić doskonałe zaplecze dydaktyczne i służyć pomocą studentom polonistyki podczas licznych zajęć kursowych. Odwołując się do klasycznego podziału mikrolingwistyki na gramatykę, semantykę i pragmatykę, dla każdego z przywołanych obszarów da się wskazać określony typ zajęć

<sup>2</sup> Wybrane przez nas, a także inne dostępne elektroniczne zasoby językowe szczegółowo omówiły w swojej publikacji Patrycja PAŁKA i Agata KWAŚNICKA-JANOWICZ (2017).

<sup>3</sup> NKJP to wspólne przedsięwzięcie Instytutu Podstaw Informatyki PAN (koordynator), Instytutu Języka Polskiego PAN, Wydawnictwa Naukowego PWN oraz Zakładu Językoznawstwa Komputerowego i Korpusowego Uniwersytetu Łódzkiego, zrealizowane jako projekt badawczy rozwojowy Ministerstwa Nauki i Szkolnictwa Wyższego w latach 2008–2012. Szczegółowemu opisowi NKJP poświęcono obszerną publikację (zob. PRZEPIÓRKOWSKI i in., red., 2012).

akademickich, podczas których można pracować z NKJP (choć zapewne w zakresie zajęć gramatycznych możliwości jest tu najmniej).

Przykładowo, w ramach kursu gramatyki opisowej swoistym uzupełnieniem ćwiczeń z zakresu przyporządkowywania wyrazów do określonych klas części mowy może być przegląd danych korpusowych za pośrednictwem wyszukiwarki Poliqarp. Zastosowana w projekcie anotacja morfoskładniowa obejmuje bowiem lematyzację segmentów, przypisanie im określonej klasy części mowy i wskazanie wartości ich kategorii gramatycznych. Student może w ten sposób połączyć utrwalanie zagadnień fleksyjnych z nauką obsługi wyszukiwarki Poliqarp, której składnia zapytań opiera się na znajomości znaczników morfosyntaktycznych. Podczas zajęć poświęconych semantyce, w szczególności semantyce leksykalnej lub leksykologii, student, przeglądając łączliwość określonych jednostek, może z kolei zrozumieć, na czym polega polisemia właściwa i systematyczna<sup>4</sup>. Obserwacje te mogłyby również być podstawą do dyskusji na temat granicy i kształtu jednostek leksykalnych. Co ważne, nawet odwołanie do klasycznych, podręcznikowych przykładów (np. *Jan kopie rów*. i *Jan kopie ziemniaki*. – poprzez wyszukanie rzeczowników pełniących funkcję dopełnienia w zdaniach z czasownikiem *kopać* w roli orzeczenia) zyskałoby na atrakcyjności w oczach „zdigitalizowanego” młodego pokolenia. Wyszucanie właściwych wniosków dotyczących polisemicznego charakteru wybranych jednostek leksykalnych to dobry punkt wyjścia do zajęć z zakresu leksykografii. Trudno dziś bowiem mówić o metodach pracy współczesnego słownikarza bez odwołań do korpusów. Stąd natomiast krok do prezentacji elektronicznych wersji leksykonów (pozostających w niniejszym opracowaniu poza obszarem naszego zainteresowania).

Biorąc pod uwagę ofertę dydaktyczną z zakresu pragmatyki językowej, praca z NKJP wydaje się szczególnie pomocna podczas zajęć poświęconych stylistyce. Wyszukiwarka PELCRA, oprócz przywołanego już kolokatora, wyposażona jest w wiele zaawansowanych funkcji, dzięki którym można między innymi ograniczyć przeglądane zbiory do określonego typu tekstów, kanału ich przekazu czy czasu publikacji. Daje to podstawę do ekscerpowania próbek według różnych parametrów, niekoniecznie równorzędnych wobec siebie (por. np. teksty literackie, internetowe a mówione). Interesująca mogłaby być analiza tekstów określonego typu (np. stenogramów sejmowych<sup>5</sup>), pozwalająca dostrzec przenikanie się różnych odmian stylistycznych. Atrakcyjne wydaje się także porównanie częstości występowania określonego słownictwa w tekstach reprezentujących różne style funkcjonalne – owe dane statystyczne można wszak wygenerować w postaci wykresów słupkowych, profilujących występowanie określonej jednostki ze względu na typ publikacji i kanał przekazu.

Oczywiście, nie sposób pominąć także przydatności NKJP w mikrobadaniach prowadzonych przez studentów podczas przygotowywania prac dyplomowych czy magisterskich. Posługując się metodą indukcyjną, dochodzą wszak do wniosków opartych na zrównoważonej próbie tekstów.

<sup>4</sup> Pomocne są tu obie wyszukiwarki korpusowe, zarówno PELCRA, a ściślej kolokator – wyszukiwarka kolokacji, jak i Poliqarp (choć w tym wypadku niezbędne byłoby zaznajomienie studentów z podstawami składni zapytań).

<sup>5</sup> Teksty te dostępne są również w odrębnej bazie danych – por. Polski Korpus Sejmowy [online: <http://sejm.nlp.ipipan.waw.pl/>; data dostępu: 12.11.2017].

## Narodowy Fotokorpus Języka Polskiego (NFJP)

Narodowy Fotokorpus Języka Polskiego (<http://nfjp.pl>) to projekt Jana Wawrzyńczyka i Piotra Wierchonia. Twórcy zaznaczają w jego opisie, że „obserwacja leksykalna objęła okres 1901–2000, zaś za kryterium wyboru przyjęto brak jednostki w pierwszym słowniku ortograficznym XX wieku – *Słowniku ortograficznym języka polskiego* Władysława Kokowskiego z roku 1903” [online: <http://nfjp.pl>; data dostępu: 19.11.2017].

Warto jednak zaznaczyć, że poza słownictwem dwudziestowiecznym w NFJP notowana jest także leksyka najnowsza, z okresu po 2001 roku (161 jednostek). Tak określone kryteria doboru tekstów czynią omawianą bazę niezwykle ciekawą. Tym, co może dodatkowo zainteresować (również studentów), jest sposób zaprezentowania poświadczeń cytatowych. Fotodokumentacja, czyli skan fragmentu tekstu, w którym została po raz pierwszy odnotowana dana jednostka leksykalna, czyni korpus atrakcyjnym i bliskim lingwistom ceniącym z jednej strony kontakt z książką tradycyjną, z drugiej – preferującym dostępność materiału w formie elektronicznej.

Mając na uwadze dydaktykę akademicką, wspomniana baza wydaje się szczególnie przydatna podczas zajęć poświęconych omawianiu koncepcji językowego obrazu świata. Przykładowo bowiem, po wpisaniu w okno wyszukiwania wyrazu *kobieta*, nie otrzymamy, co jasne, poświadczenia użycia tego właśnie słowa, a listę wyników złożoną z 258 jednostek – wyrazów (prefigowanych, dla których bazą jest rzeczownik *kobieta*) lub zestawień i złożzeń z członem *kobieta*. Wśród wyników wyszukiwania znajdują się zarówno takie, które możemy uznać za stosunkowo częste (np. *superkobieta/super-kobieta*, *kobieta-chirurg*, *kobieta-szpieg*, *kobieta-demon*), jak i użycia wyjątkowe (np. *zookobieta*, *metakobieta*, *kobieta-pszczoła*, *kobieta-nietoperz*). Daje to podstawę do dyskusji na temat sposobów postrzegania kobiety przez użytkowników polszczyzny w ostatnim stuleciu. Bardzo praktyczne są w tych poszukiwaniach indeksy *a fronte* i *a tergo*, wyświetlające się podczas oglądania konkretnego wyniku, wskazujące sąsiedztwo wyszukiwanej jednostki pośród innych odnotowanych w bazie. Biorąc pod uwagę fakt, że NFJP skupia w większości derywaty lub neologizmy słowotwórcze, można również wykorzystać jego zasoby do dyskusji nad tendencjami słowotwórczymi czy szerzej: rozwojowymi polszczyzny podczas różnych zajęć poświęconych współczesnemu językowi polskiemu.

## ChronoPress – Portal Tekstów Prasowych

Prezentując dostępne w Internecie korpusy językowe, nie sposób pominąć bazy, jaką jest ChronoPress (<http://chronopress.clarin-pl.eu>) – korpus i wyszukiwarka, które powstały w ramach prac konsorcjum CLARIN-PL. Korpus zawiera około 56 tysięcy fragmentów tekstów prasowych opublikowanych w latach 1945–1954. Jak zaznaczają w opisie bazy jej twórcy, próbki „reprezentują pełne spektrum tematyczne oficjalnego dyskursu publicznego” wskazanego okresu [online: <http://chronopress.clarin-pl.eu>; data dostępu: 20.11.2017]. Z tego względu mogą one okazać się świetnymi ilustracjami podczas omawiania ze studentami przemian w zakresie kształtowania się stylu publicystycznego w XX wieku – zarówno jego warstwy leksykalnej, jak i cech składniowych.

Narzędzie to oferuje kilka możliwości. Zakładka Konkordacje pozwala wyekscerpować określone jednostki leksykalne w kontekście (ograniczonym do jednego zdania), a opcja Szeregi czasowe – generować wykresy, ukazujące w skali rocznej lub miesięcznej częstotliwość użycia danej jednostki w tekstach stanowiących korpus. Wyszukując słownictwo związane z realiami historycznymi lub kulturalnymi, można zatem obserwować zakres jego użycia na przełomie lat 40. i 50. XX wieku. Być może cenne okazałoby się porównanie tych wyników ze stanem współczesnym, a zatem połączenie pracy z ChronoPressem i na przykład NKJP (z ograniczeniem do tekstów publicystycznych).

Jedną z funkcji ChronoPressu jest również Mapa nazw miejscowych, na której zostały zaznaczone wszystkie nazwy miast i wsi odnotowane w tekstach korpusu. Ich zbiór, poszerzony o dostępne online kartoteki nazw miejscowych, stanowi bogatą bazę do rozważań z zakresu onomastyki, a także do obserwacji zmian postaci ortograficznej wybranych nazw własnych (w stosunku do formy dzisiejszej lub przeciwnie, do dawnej, wyekscerpowanej ze źródeł leksykograficznych).

Omawiany korpus mógłby zatem posłużyć jako atrakcyjne uzupełnienie zajęć kursowych, podczas których dokonuje się (z różnych perspektyw) oglądu polszczyzny z połowy XX wieku.

### Wyszukiwarka danych konwersacyjnych Spokes

W ramach działań konsorcjum naukowego CLARIN-PL powstało wiele elektronicznych zasobów językowych oraz aplikacji, które są „przeznaczone do konkretnych zadań badawczych w naukach humanistycznych i społecznych oraz narzędzi do automatycznej analizy tekstu w języku polskim” [online: <http://clarin-pl.eu/pl/uslugi/>; data dostępu: 15.11.2017]. Jednym z rozwiązań tego typu, oprócz omówionego uprzednio ChronoPressu, jest Spokes (<https://spokes.clarin-pl.eu>) – platforma internetowa stworzona z inicjatywy Piotra Pęczika z Uniwersytetu Łódzkiego, na którą składają się dwa komplementarne elementy: korpus języka mówionego oraz wyszukiwarka danych konwersacyjnych. Takie przedsięwzięcie stanowi przykład elektronicznego zasobu językowego polszczyzny mówionej, tzw. korpusu konwersacyjnego, zawierającego obecnie ponad 240 tysięcy transkrybowanych wypowiedzi (tj. ok. 2,4 miliona słów) oraz ich odpowiedników w formie plików dźwiękowych (plików audio). Na zbiór ten składają się przede wszystkim swobodne i spontaniczne rozmowy codzienne (576 rozmów, głównie *in vivo* z lat 2000–2015), nagrane z ukrycia przez ekspertów.

W perspektywie szerokiego spojrzenia na zastosowanie aplikacji Spokes w dydaktyce akademickiej należy podkreślić, że stanowi ona przykład cennego zasobu i zarazem narzędzia do badań nad współczesną polszczyzną mówioną. Opcja zarówno bezpośredniego odtworzenia, jak i pobrania wspomnianych plików audio, które umieszczone są przy każdym wyniku wyszukiwania, stwarza szansę wykorzystania tego typu materiału jako uzupełnienia zajęć z fonetyki współczesnego języka polskiego – wówczas wyekscerpowane przykłady mogą stać się podstawą do zilustrowania różnych procesów fonetycznych czy też kwestii związanych z zagadnieniami akcentu i intonacji.

Filtrowanie materiału językowego według zmiennych socjolingwistycznych, takich jak wiek, płeć czy wykształcenie rozmówców, okazuje się przydatne do badań nad językiem określonych grup społecznych między innymi w ramach zajęć z socjolingwistyki. Przykładowo, można poddać analizie słownictwo młodzieży, by obserwować częstotliwość występowania określonych zapożyczeń, neologizmów lub potoczizmów (zestawiając te wyniki dla porównania z wypowiedziami osób starszych).

Ponieważ głównym założeniem twórców platformy Spokes jest gromadzenie zasobów konwersacyjnych i możliwość ich zaawansowanego przeszukiwania, narzędzie to znajdzie zastosowanie w rozważaniach z zakresu szeroko pojętej pragmatyki, w tym nad strukturą dialogu, regułami konwersacyjnymi, etykietą językową czy w badaniach dyskursu mówionego. Interesujące mogłoby okazać się wykorzystanie nagranych rozmów codziennych w celu rozpoznawania różnych typów aktów mowy lub dostrzegania w poszczególnych wypowiedziach rozbieżności między celem illokucyjnym a efektem perlokucyjnym. Atrakcyjne wydaje się także, szczególnie w kontekście charakterystyki zebranego materiału, wyzyskanie nagrań do zilustrowania korzyści wynikających z pogwałcenia norm konwersacyjnych.

### Wyszukiwarka korpusowa Monco PL

Ciekawą propozycją jest również wyszukiwarka korpusowa Monco PL (<http://monco.frazeo.pl/>). Przeszukiwana z jej użyciem baza leksykalna, licząca około 2,5 miliarda słów, jest codziennie aktualizowana, składa się bowiem wyłącznie z materiałów internetowych, pozyskiwanych między innymi z wiodących portali o różnym profilu tematycznym (np. onet.pl, interia.pl, naszemiasto.pl, kulturaonline.pl, sejm.gov.pl, egospodarka.pl, plotek.pl) czy stron prowadzonych przez wydawców prasowych, telewizyjnych i radiowych (np. gazeta.pl, fakt.pl, tvn24.pl, tvp.info, rmf24.pl). Monco PL nie jest więc korpusem, a narzędziem, które opierając się na składni zapytań, właściwiej wyszukiwarce PELCRA, obsługującej NKJP, pozwala przeglądać zasoby internetowe wyselekcjonowane i ograniczone do tysiąca źródeł (pod względem stylistycznym są to w dużej mierze teksty publicystyczne). Daje tym samym szansę odnalezienia przykładów użycia słownictwa najnowszego, co można uznać za swoiste uzupełnienie NKJP. I właśnie ze względu na zasięg czasowy tekstów przeszukiwanych za pomocą omawianego narzędzia jest ono szczególnie pomocne podczas omawiania w trakcie zajęć ze studentami tendencji rozwojowych współczesnej polszczyzny. Atrakcyjna wydaje się w tym celu funkcja, znajdująca się w zakładce Fasety, która prezentuje wykresy określające procentową obecność szukanego słowa w dziesięciu najpopularniejszych źródłach. Dwa kolejne diagramy wskazują z kolei dzienne częstości wystąpienia danego wyniku z ostatnich trzech miesięcy oraz tygodniowe z ostatniego roku. Uzyskane w ten sposób dane statystyczne mogą zostać wykorzystane podczas zajęć ze stylistyki – student, dokonując ekscerpcji wybranego słownictwa, dowodnie przekonuje się o kontekstach jego użycia oraz typach tekstów, w których ono występuje. Przeglądając natomiast obszerniejsze próbki wypowiedzi, student może wnioskować przykładowo na temat cech składniowych czy gatunkowych tekstów reprezentujących określony styl.

Niewątpliwie wyszukiwarka Monco PL służy również jako pomoc podczas gromadzenia materiału na potrzeby pisanie pracy dyplomowej czy magisterskiej odnoszącej się do najnowszej leksyki – stanowi bowiem lepsze pod względem badawczym źródło niż na przykład wyszukiwarka internetowa Google.

### Korpusomat – narzędzie do tworzenia przeszukiwalnych korpusów języka polskiego

Omówione dotychczas wyszukiwarki korpusowe pozostają w ścisłym związku z określonymi bazami danych językowych. Nieocenionym narzędziem w pracy językoznawcy, a także inspiracją do działań dydaktycznych, może okazać się również własny elektroniczny korpus tekstów. Przykładem aplikacji, która pozwala na opracowanie przeszukiwalnych baz tekstów na podstawie zaimportowanych przez użytkownika plików, jest Korpusomat (<http://korpusomat.nlp.ipipan.waw.pl/>) – serwis internetowy opracowany przez zespół z Instytutu Podstaw Informatyki PAN. Wybrane do korpusu teksty są automatycznie poddawane konwersji kodowania tekstu do UTF-8 oraz analizie morfologicznej (za pomocą analizatora morfologicznego Morfeusz i *Słownika gramatycznego języka polskiego*), a także oznakowane morfosyntaktycznie (za pomocą tagera Concraft), by ostatecznie mogły zostać skompilowane w postaci plików binarnych. Tak utworzony korpus można przeszukiwać za pomocą wyszukiwarki Poliqarp, znanej choćby z NKJP. Korpusomat pozwala więc tworzyć dowolne bazy tekstów według rozmaitych kryteriów – tematu, dziedziny, autora, epoki, odmian języka itd. Narzędzie to umożliwi także, co wydaje się szczególnie interesujące w odniesieniu do badań nad współczesną polszczyzną, zwłaszcza nad językiem mediów, pobieranie tekstów ze wskazanych stron internetowych lub innych źródeł elektronicznych.

Przykładem wykorzystania Korpusomatu w dydaktyce akademickiej mogą być również ćwiczenia z zakresu stylistyki, podczas których analiza korpusowa mogłaby obejmować obliczanie częstotliwości występowania wybranych słów, fraz i kolokacji w tekstach reprezentujących różne style funkcjonalne polszczyzny, na przykład styl naukowy, publicystyczny, religijny czy potoczny. Interesujące z punktu widzenia analiz stylistycznych dotyczące odmiany artystycznej i jej gatunków, a także idiolektu autora, okazuje się badanie najczęstszych kontekstów, w których występują dane słowa lub zwroty. To tylko wybrane z wielu propozycji ćwiczeń, jakie mogą być zrealizowane z wykorzystaniem Korpusomatu. Przede wszystkim jednak, podobnie jak inne aplikacje będące przedmiotem oglądu w niniejszym artykule, narzędzie to może usprawnić pracę własną studenta podczas przygotowania prac pisemnych czy zaliczeniowych, których zadaniem jest analiza samodzielnie zebranego materiału językowego.

### Wirtualne Laboratorium Transkrypcji – aplikacja na potrzeby badań diachronicznych

Materiał do badań językowych można pozyskiwać na wiele sposobów. Poza tradycyjnymi metodami – kwerendami w bibliotekach i archiwach – coraz częściej sięga się po źródła

elektroniczne w postaci zdigitalizowanych wydawnictw (w przypadku dawnych źródeł są to skany lub zdjęcia zbiorów). Także wizyta w bibliotece pozwala samodzielnie zeskanować lub sfotografować interesującą nas publikację. Tak skopiowany materiał źródłowy nie umożliwia jednak edycji ani przeszukiwania, co znacznie utrudnia pracę badawczą (nie można np. wykorzystać go do stworzenia bazy za pomocą wcześniej omówionego narzędzia, jakim jest Korpusomat). Przykładem rozwiązania, które wspomaga proces analizy zebranych danych, jest narzędzie przygotowane przez Poznańskie Centrum Superkomputerowo-Sieciowe (PCSS) – Wirtualne Laboratorium Transkrypcji (<http://wlt.pcss.pl/wlt-web/index.xhtml>). Pozwala ono tworzyć cyfrowy i przeszukiwalny tekst na podstawie zeskanowanych dokumentów, w tym także dokumentów historycznych.

Wiedza o historycznym rozwoju polszczyzny stanowi podstawę zrozumienia współczesnego stanu języka, dlatego tak potrzebne w dydaktyce akademickiej jest nieustanne urozmaicanie sposobów wprowadzania zagadnień historycznojęzykowych. Nierzadko od atrakcyjności tego przekazu zależy zainteresowanie nim studentów. Z pomocą przychodzi tu Wirtualne Laboratorium Transkrypcji, będące bardzo ciekawym narzędziem do pracy podczas zajęć z historii języka. Student może samodzielnie zdigitalizować fragment na przykład dawnej kroniki za pomocą opcji OCR<sup>6</sup> oraz, edytując zeskanowany tekst, zweryfikować poprawność wygenerowanego zapisu na podstawie wiedzy na temat zjawisk z zakresu fonetyki i fleksji historycznej. Przeprowadzona może też zostać analiza tekstów dawnych ze względu na określone kryterium (np. obserwacja i określanie etapów rozwoju liczebników). Oglądowi poddać można dowolne źródło historyczne, na przykład fragmenty *Nowego Testamentu* w przekładzie Jakuba Wujka, w celu zwrócenia uwagi na obecne tam między innymi szesnastowieczne formy czasu przeszłego.

Z punktu widzenia nauczania treści historycznojęzykowych interesującym narzędziem, które wchodzi w skład Wirtualnego Laboratorium Transkrypcji, jest aplikacja Wycinanki. Służy ona do precyzyjnego opracowania tekstów dawnych i wspomaga proces automatycznego rozpoznawania znaków. W ramach pracy z aplikacją należy szczegółowo przyjrzeć się materiałowi źródłowemu i samodzielnie poprawić te znaki diakrytyczne, które zostały błędnie odczytane przez OCR. Możliwość szczegółowego oglądu treści i przede wszystkim zapisu dawnych dokumentów da się wykorzystać do zilustrowania historii kształtowania się polskiej grafii. Studenci, edytując fragmenty wybranych dokumentów, mogą obserwować sposób zapisu iloczasu samogłosek ustnych czy też rozwiązania dotyczące rozróżnienia samogłosek jasnych i pochylonych. Możliwości wykorzystania Wirtualnego Laboratorium Transkrypcji w dydaktyce akademickiej jest znacznie więcej – począwszy od drobnych ćwiczeń gramatycznych czy leksykalnych, aż po badania własne studentów na potrzeby przygotowywania prac dyplomowych i magisterskich. Niewątpliwie jest to aplikacja, która pozwoli uatrakcyjnić zajęcia kursowe i ukazać potencjał badań historycznojęzykowych.

---

<sup>6</sup> Skrót OCR pochodzi od angielskiego Optical Character Recognition, czyli optycznego rozpoznawania znaków na skanie dokumentów.



## Słowosieć – pierwszy polski wordnet

Zastosowanie w dydaktyce akademickiej znajdują nie tylko narzędzia związane z bazami tekstów. Interesującą propozycję konsorcjum CLARIN-PL stanowi Słowosieć (<http://plwordnet.pwr.wroc.pl/wordnet/>) – leksykalna sieć semantyczna języka polskiego opracowana przez zespół z Grupy Technologii Językowych Politechniki Wrocławskiej. Jest to pierwszy polski wordnet wzorowany na platformie Princeton WordNet dla języka angielskiego; oba projekty zostały połączone, dzięki czemu mogą służyć także jako słownik polsko-angielski i angielsko-polski. Obecnie wielka sieć wyrazów liczy ponad 178 tysięcy słów, 259 tysięcy znaczeń, ponad 600 tysięcy relacji oraz około 158 tysięcy haseł polsko-angielskich i według deklaracji jej twórców stanowi największy tego typu zbiór na świecie. Założeniem projektu jest nieustanny rozwój i tworzenie kolejnych zestawień znaczeń połączonych wzajemnymi relacjami, gdzie „każdy wyraz jest zdefiniowany poprzez odniesienie do innych wyrazów” [online: <http://plwordnet.pwr.wroc.pl/wordnet/about>; data dostępu: 14.10.2017]. Choćby z tego powodu narzędzie to okazuje się bardzo przydatne na zajęciach z semantyki lub leksykologii. Słowosieć można bowiem wykorzystać do ćwiczeń służących ukazaniu relacji, takich jak: hipo-/ hiperonimia, meronimia, holonimia, synonimia, antonimia, fuzynimie czy konwersja. Przykładowo, student ma możliwość wyszukiwania jednostek leksykalnych, tworzenia z nich ciągów synonimicznych i określania różnic znaczeniowych między nimi (dotyczących zakresu użycia, stopnia intensywności wyrażanej cechy itd.). Obserwacja sieci powiązań semantycznych mogłaby okazać się dla studentów polonistyki cennym ćwiczeniem utrwalającym poznane w trakcie zajęć treści.

Słowosieć to atrakcyjne dla słuchaczy narzędzie zarówno ze względu na możliwość samodzielnego wyszukiwania leksemów i tym samym praktycznego zastosowania strategii cyfrowych, takich jak BYOD<sup>7</sup> (szczególnie z wykorzystaniem urządzeń mobilnych, ponieważ opisywany wordnet jest także dostępny w formie aplikacji), jak i na potencjał prezentowania informacji w formie wizualnej, czyli w postaci edytowalnych grafów, które ilustrują sieć powiązań między poszczególnymi jednostkami języka. Dzięki temu studentom łatwiej zapamiętać relacje znaczeniowe między wyrazami, ponieważ, zgodnie z założeniami kształcenia akademickiego, wizualizacja jest cennym „środkiem wspomaganie podmiotowego tworzenia wiedzy i gromadzenia jej w kodach hierarchicznych” (DYLAK, 1995: 9).

## Jasnopis – narzędzie do mierzenia zrozumiałości tekstu

Ostatnim z proponowanych przez nas narzędzi, możliwych do wykorzystania w dydaktyce akademickiej, jest Jasnopis (<http://www.jasnopis.pl/>). Ten program informatyczny, za którego pomocą możliwe jest zmierzenie poziomu zrozumiałości wypowiedzi, rozwiązuje problem istnienia w przestrzeni publicznej zbyt trudnych tekstów, do których należą między innymi pisma urzędowe, instrukcje, akty prawne. Opracowana autorska formuła analityczna obliczająca zrozumiałość powstała na podstawie złożonych badań psycholin-

<sup>7</sup> BYOD to akronim od ang. *Bring Your Own Device* ('przyjdź z własnym sprzętem').

gwistycznych, statystycznych oraz metod korpusowych<sup>8</sup>, a także odniesień do znanych formuł/wzorów – indeksu trudności tekstu polskiego Walerego Pisarka oraz wskaźnika mglistości tekstu (FOG)<sup>9</sup>.

Narzędzie adresowane jest przede wszystkim do osób, które tworzą teksty użytkowe, ale także do badaczy zajmujących się czytelnością i jasnością wypowiedzi, co stwarza możliwości wykorzystania Jasnopisu również na zajęciach kursowych skierowanych do studentów polonistyki. Znajdzie on zastosowanie w trakcie ćwiczeń poświęconych pragmatyce językowej, podczas których poddawane badaniom teksty stanowić mogą materiał do rozważań z zakresu stylistyki. Za sprawą złożonych parametrów analizy treści narzędzie to pozwala bowiem na wyodrębnienie wielu zmiennych językowych i ocenę cech stylistycznych danego tekstu. Do wspomnianych wyróżników należą między innymi stosunek liczby rzeczowników do czasowników, średnia długość zdania, długość słowa, średnia długość akapitu w słowach czy procent słów trudnych. Przykładowo, interesujące i zachęcające do pracy własnej mogłyby być dla studentów ćwiczenia dotyczące wskazywania cechy stylistycznej danego tekstu: nominalności lub werbalności stylu.

W ramach zajęć z zakresu genologii lingwistycznej studenci mogą z kolei za pomocą Jasnopisu poddawać oglądowi fragmenty tekstów użytkowych (np. urzędowych, prawniczych, naukowych) i podejmować próby ustalenia ich wzorca gatunkowego. Jest to możliwe dzięki zaawansowanym ustawieniom narzędzia – grafu podobieństwa (który porównuje daną wypowiedź z korpusami referencyjnymi zawierającymi m.in. teksty popularnonaukowe, ustawy czy artykuły prasowe) oraz trybu testów Taylora (dzięki któremu wyliczane jest podobieństwo języka (stylu) tekstu do jednego z kilku modeli językowych).

Jasnopis stwarza bardzo duże możliwości także w zakresie opracowywania tekstów – pisemnych prac zaliczeniowych czy też prac dyplomowych. Narzędzie dostępne jest w formie wtyczki do edytorów tekstu (Microsoft Word i OpenOffice), co pozwala na tworzenie czytelnych w odbiorze wypowiedzi pisemnych już w trakcie bezpośredniej pracy nad nimi – kontrola poziomu trudności tekstu odbywać się może na bieżąco. Tekst przyporządkowany zostaje do jednej z siedmiu klas trudności, a program wyróżnia w nim miejsca wymagające korekty, ponieważ zostały uznane za mało czytelne dla odbiorcy. Spośród typów wyróżnień, które mogą okazać się pomocne także w zakresie poszerzenia wiedzy na temat relacji semantycznych oraz zwiększenia zasobu leksykalnego, należy wspomnieć o wyświetlaniu podpowiedzi dotyczących zamiany zbyt trudnych wyrazów na ich łatwiejsze odpowiedniki – Jasnopis prezentuje proponowane synonimy, hiponimy i/lub hiperonimy na podstawie informacji pozyskanych z omawianego już wcześniej narzędzia – Słowosieci.

Dokonany przegląd przede wszystkim wskazuje możliwości wykorzystania zasobów cyfrowych podczas zajęć poświęconych współczesnemu językowi polskiemu w różnych jego odsłonach, głównie jednak na płaszczyźnie semantyki i stylistyki. Nie oznacza to, że diachroniści mają do dyspozycji wyłącznie Wirtualne Laboratorium Transkrypcji. Z inicjatywy zespołu lingwistek pracujących w Instytucie Języka Polskiego im. Ireny Bajerowej

---

<sup>8</sup> Więcej na ten temat można znaleźć w publikacji, której autorami są twórcy omawianego narzędzia: GRUSZCZYŃSKI, OGRODNICZUK, red., 2015.

<sup>9</sup> Patrz [online: [www.jasnopis.pl/static/pdf/jasnopis-wprowadzenie.pdf](http://www.jasnopis.pl/static/pdf/jasnopis-wprowadzenie.pdf); data dostępu: 29.11.2017].

na Uniwersytecie Śląskim w Katowicach utworzona została platforma internetowa [www.korpusydiachroniczne.pl](http://www.korpusydiachroniczne.pl), która służy upowszechnianiu dziedzictwa kulturowego i naukowego polszczyzny w ujęciu diachronicznym. Docelowo ma ona stanowić kompendium wiedzy na temat istniejących oraz powstających korpusów diachronicznych i inspirację nie tylko dla badaczy, ale również dla dydaktyków i studentów.

Z pewnością możliwości wykorzystania zasobów cyfrowych, również tych omówionych w artykule, jest znacznie więcej. Najistotniejsza wydaje się jednak świadomość, że w dzisiejszych realiach informatyczne zdolności studentów i ich „cyfrowe zaangażowanie” nie są czymś, co odciąga od nauki, lecz potencjałem, z którego należy korzystać w procesie dydaktycznym, także w świecie humanistów.

## Słowniki

SALONI Z., WOLIŃSKI M., WOŁOSZ R., GRUSZCZYŃSKI W., SKOWROŃSKA D., 2015: *Słownik gramatyczny języka polskiego*. Warszawa.

## Źródła

ChronoPress – Portal Tekstów Prasowych [online: <http://chronopress.clarin-pl.eu/#!start>; data dostępu: 20.11.2017].

Jasnopis [online: [www.jasnopis.pl/](http://www.jasnopis.pl/); data dostępu: 29.11.2017].

Korpusomat [online: <http://korpusomat.nlp.ipipan.waw.pl/>; data dostępu: 27.11.2017].

Korpusy diachroniczne polszczyzny [online: [www.korpusydiachroniczne.pl](http://www.korpusydiachroniczne.pl/); data dostępu: 29.11.2017].

Monco PL [online: <http://monco.frazeo.pl/>; data dostępu: 20.11.2017].

NFJP: Narodowy Fotokorpus Języka Polskiego [online: <http://nfjp.pl/>; data dostępu: 19.11.2017].

NKJP: Narodowy Korpus Języka Polskiego [online: <http://nkjp.pl/>; data dostępu: 12.11.2017].

Polski Korpus Sejmowy [online: <http://sejm.nlp.ipipan.waw.pl/>; data dostępu: 12.11.2017].

Słowosieć [online: <http://plwordnet.pwr.wroc.pl/wordnet/>; data dostępu: 14.10.2017].

Wirtualne Laboratorium Transkrypcji [online: <http://wlt.synat.pcass.pl/wlt-web/index.xhtml>; data dostępu: 20.10.2017].

Wyszukiwarka danych konwersacyjnych Spokes [online: <https://spokes.clarin-pl.eu/>; data dostępu: 15.11.2017].

## Literatura

BIŁAS-PLESZAK E., SUJKOWSKA-SOBISZ K., 2016: *Uniwersytet wobec zmian i potrzeb współczesności. Kilka uwag na temat kierunków studiów o profilu praktycznym*. „Forum Lingwistyczne”, nr 3, s. 167–172.

BOMBA R., 2013: *Narzędzia cyfrowe jako wyznacznik nowego paradygmatu badań humanistycznych*. W: RADOMSKI A., BOMBA R., red.: *Zwrot cyfrowy w humanistyce*. Lublin, s. 57–71.

ChronoPress – Portal Tekstów Prasowych [online: <http://chronopress.clarin-pl.eu/#!start>; data dostępu: 20.11.2017].

CLARIN-PL [online: <https://clarin-pl.eu/pl/uslugi/>; data dostępu: 15.11.2017].

- DigiLing – Trans-European eLearning Hub for Digital Linguistics [online: <http://www.digiling.eu/>; data dostępu: 21.11.2017].
- DYLAK S., 1995: *Wizualizacja w kształceniu nauczycieli*. Poznań.
- GRUSZCZYŃSKI W., OGRÓDNICZUK M., red., 2015: *Jasnopis, czyli mierzenie zrozumiałości polskich tekstów użytkowych*. Warszawa.
- NFJP: Narodowy Fotokorpus Języka Polskiego [online: <http://nfjp.pl/>; data dostępu: 19.11.2017].
- PAŁKA P., KWAŚNICKA-JANOWICZ A., 2017: *Przewodnik po elektronicznych zasobach językowych dla polonistów (słowniki, kartoteki, korpusy, kompendia)*. Kraków [online: [http://tmjp.pl/images/pdf/przewodnik\\_po\\_elektronicznych\\_zasobach.pdf](http://tmjp.pl/images/pdf/przewodnik_po_elektronicznych_zasobach.pdf); data dostępu: 29.11.2017].
- PRZEPIÓRKOWSKI A., BAŃKO M., GÓRSKI R., LEWANDOWSKA-TOMASZCZYK B., red., 2012: *Narodowy Korpus Języka Polskiego*. Warszawa.
- SAJDUK B., 2014: *Nowoczesna dydaktyka akademicka. Kto kogo uczy?* Kraków [online: <http://dydaktyka-akademicka.pl/>; data dostępu: 29.11.2017].
- Słowność [online: <http://plwordnet.pwr.wroc.pl/wordnet/>; data dostępu: 14.10.2017].

Beata Duda, Karolina Lisczyk

Digital Tools in the Polish Language Academic Didactics –  
Applications, Possibilities, Prospects

#### Summary

The aim of the article is to present the selected digital tools for language analysis which may be applied in the academic didactics during course classes in linguistics. In the first part of the article advanced corpus search engines were discussed, for instance, National Corpus of Polish (NKJP), National Photo-corpus of the Polish Language (NFJP), as well as ChronoPress – Press Texts Portal, Spokes – corpora of conversational speech, and corpus search engine Monco PL. The second part is dedicated to the tools and applications for the individual creation of browsable language resources (Korpusomat, Virtual Laboratory of Transcription), as well as for the complex analysis of the relationships between language units (plWordnet) and the measurement of the degree of the text comprehension (Jasnopis).

Key words: digital humanities, digital linguistics, digital tools, academic didactics